DOCUMENT RESUME

ED 242 768 TM 840 182

AUTHOR	Pfau, Richard Henry
TITLE	Standardizing Behavioral Measurements across Cultures, Nations, and Time. World Education Monograph Series: Number One, 1983-1984.
INSTITUTION	Connecticut Univ., Storrs. Thut (I.N.) World Education Center.
REPORT NO	ISBN-0-918158-27
PUB DATE	[84]
NOTE	38p.; For related document, see ED 198 178.
AVAILABLE FROM	I.N. Thut World Education Center, Box U-32, School of Education, The University of Connecticut, Storrs, CT, 06268 ($1.50 plus $0.30 postage).
PUB TYPE	Reports - Research/Technical (143) -- Viewpoints (120)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Behavior Rating Scales; *Cross Cultural Studies; Evaluation Methods; International Studies; *Measurement Techniques; *Observation; Research Methodology; *Research Needs; *Standards

ABSTRACT

This monograph reviews studies which have measured and compared classroom and other human behaviors occurring in different cultures and nations, points out problems related to the comparisons made, and describes procedures which can be used to help standardize measurements of behavior made using systematic observation instruments. Standardization is considered to be achieved when behaviors are classified the same way by different observers who use an instrument, and when measurements which result have scalar identity and are free of systematic observation errors. Procedures discussed include using observers from each culture studied, and preserving instrument descriptions, samples of behaviors studied, and associated standard measurements of those behaviors for reference by others. Areas needing additional research and thought are also highlighted. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED242768

ISBN 0-918158-27

STANDARDIZING BEHAVIORAL MEASUREMENTS ACROSS CULTURES, NATIONS, AND TIME

by

Richard Henry Pfau, Ph.D.
Associate Extension Professor
Institute of Public Service, International
The University of Connecticut
Hartford, Ct.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

F Stone

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

1983-1984
WORLD EDUCATION MONOGRAPH SERIES
Number One

THE I.N. THUT WORLD EDUCATION CENTER

Centro de Educacion Mundial I.N. Thut
Box U-32, School of Education
The University of Connecticut
Storrs, Connecticut 06268

TM 840 182

FOREWORD

Since 1977 The I.N. Thut World Education Center (TWEC) has issued World Education Monographs. They are usually the products of student or faculty research, or the texts of presentations that were made at conferences or colloquia. A complete list of the sixteen monographs that have been issued to date is available from TWEC.

The 1982-1983 titles in the series were:

AL-AZHAR: A UNIVERSITY BETWEEN TWO AGES.
Mohamed Misbah, 1983. 13 pp.

THE EGYPTIAN EXPERIENCE IN EDUCATION.
Mohamed Misbah, 1983. 20 pp.

CULTURAL CONFLICTS IN CARRYING OUT A CHILD FEEDING PROJECT IN PARAGUAY: THE ITACURUBI DEL ROSARIO EXPERIENCE OF BETTY KEENEY, 1945-1947.
(Double Issue) Kay Hill, 1983. 60 pp.

Copies of these publications may be obtained by sending a check for $1.50 (except for the last title which costs $3.00) made out to The I.N. Thut World Education Center for each monograph ordered. Please add $1 to cover postage and handling for orders of less than three items, or $2 for more than three. All orders must be pre-paid or on official purchase forms. Discounts are not offered to book sellers or purchasing agents, and TWEC cannot accept returns or make refunds.

In addition to the present monograph, the 1983-1984 series is scheduled to include:

CONTEMPORARY EDUCATION IN GHANA, WEST AFRICA.
Ellen Segbefia

COMPARING PUBLIC AND PRIVATE SCHOOLS IN GHANA: THE PUZZLING ROLE OF SELECTIVITY BIAS.
Bernard Kodwo Hayford

THE INFLUENCE OF ECONOMIC AND OTHER FACTORS ON THE QUALITY OF THE TEACHER WORKFORCE: A CASE STUDY OF GHANA.
Bernard Kodwo Hayford

The author of the present monograph, Dr. Richard H. Pfau, is a native of Baltimore, Maryland. He received his undergraduate degree from the University of Baltimore, and earned his doctorate in International Development Education at the University of Pittsburgh. Dr. Pfau worked for more than seven years in Nepal as a member of the Peace Corps, and under AID auspices. He is married, and the Pfaus have two children. They now make their home in Mansfield Center, Connecticut.

ABSTRACT

This monograph reviews studies which have measured and compared classroom and other human behaviors occurring in different cultures and nations, points out problems related to the comparisons made, and describes procedures which can be used to help standardize measurements of behavior made using systematic observation instruments. Standardization is considered to be achieved when behaviors are classified the same way by different observers who use an instrument, and when measurements which result have scalar identity and are free of systematic observation errors. Procedures discussed include using observers from each culture studied, and preserving instrument descriptions, samples of behaviors studied, and associated standard measurements of those behaviors for reference by others. Areas needing additional research and thought are also highlighted.

TABLE OF CONTENTS

	Page
Introduction	1
Category Systems	2
Standardization	5
Past Usage of Category Systems	7
Limitations of Past Category System Usage	10
A Suggested Approach	16
Some Needed Research and Thinking	19
Footnotes	22
References	24

Standardizing Behavioral Measurements Across Cultures, Nations, and Time

Standardized behavioral measurements are lacking in the social sciences. Although institutes, associations, and hundreds of standards have been established to facilitate the making of physical measurements for engineering, physical science, and business purposes, virtually no mechanisms or accepted standards exist to help social scientists make more comparable measurements of behavior.[1] As a result, studies of behavior often lack precision and validity, scholars are hampered in their ability to communicate with one another, and the development and testing of social science theory is hindered (Moles, 1977; Triandis, 1977, p. 10; Johnson, 1978; Nunnally, 1978, pp. 6-10).

This monograph explains procedures which can be used to help standardize measurements of behavior made using systematic observation techniques, discusses problems encountered when such techniques are used to measure and compare naturally occurring behaviors across cultures and time, and indicates areas of inquiry which, if pursued, could provide information useful to comparative scholars wishing to standardize behavioral measurements.

Category Systems

Behavior can be observed and measured in many ways. Relatively indirect ways include the use of questionnaires, interviews, and diaries to obtain information from persons about their own behavior or the behavior of others. More direct ways include having trained persons directly observe behaviors of interest and record their observations by writing narrative descriptions, or by means of rating systems, checklists, or other observation instruments.

One systematic technique which provides an especially promising basis for making precise and valid cross-cultural comparisons of behavior involves the use of "category systems" (Pfau, 1976, 1980). This is the technique upon which the present discussion is focused.

Category systems are systematic observation instruments which are characterized by two major features: (a) clearly specified, well defined categories of behavior to be measured, and (b) objective means for recording the occurrence of those behaviors, such as counting methods or the use of timing devices (as indicated in Table 1). Observers using these instruments make records of behaviors observed as those behaviors occur or within a few seconds afterwards. Alternatively, records may be made at a later time, by viewing films, videotapes or other preserved samples of behaviors to be measured.

TABLE 1

CATEGORY SYSTEM RECORDING METHODS

Recording method[a]	Distinguishing characteristics	Accuracy
Tally method frequency recording event recording	A record is made each time a behavior of interest occurs.	Potentially high
Duration method	A cumulative stopwatch or other timing device is started when the behavior of interest begins and is stopped when the behavior ends. Alternatively, the beginning and ending times of behaviors are recorded on paper or on a special recording instrument.	Potentially high
Instantaneous time sampling method	Records are made of the behaviors occurring at exact instants of time. These instants are often separated by fixed periods of time, such as 30 seconds or 5 minutes.	Potentially high
Interval method whole interval time sampling partial-interval time sampling one-zero sampling	The observation period is divided into small intervals of time, lasting from 3 seconds to 15 seconds or more. Recordings are made to indicate whether behaviors of interest were observed to occur during each time interval.	Variable[b]

TABLE 1 (Continued)

Recording method[a]	Distinguishing characteristics	Accuracy
Checklist[c]	Same as the "interval method" but with a relatively larger time interval.	Variable (Generally lower than other methods listed above)
Specimen record[c] ad libitum sampling	A detailed narrative or shorthand description is made of behavior as it is observed. Later, the occurrence of specific behaviors of interest are counted or otherwise classified.	Apparently lower than most other methods described above

[a]The first four major classifications shown are based on Jackson, Della-Piana, and Sloane (1975). Alternative names and slight variations of the major methods are also indicated. See Jackson et al. (1975) and Altmann (1974) for details of the first four methods described.

[b]The interval method should be used with caution. It is subject to differentially distorting measurements of behaviors observed in different cultures.

[c]These are variations of category systems.

Often, observers who use these instruments are trained until their records agree highly with those made by experts in the use of an instrument. High agreement, when reached, indicates that the process of measurement is unambiguous, and that a "standard language" has been applied by the different observers to describe behaviors observed.

The potential suitability of category systems for helping to make cross-cultural comparisons of behavior is a result of the explicit classification and objective recording procedures associated with use of these instruments. These procedures lend themselves to being used by observers of differing cultural backgrounds to make standardized measurements of behavior. Such standardized measurements, in turn, provide a basis for making precise and valid cross-cultural comparisons (see Pfau, 1980, for details).

Standardization

Before going further, let me explain what I mean by standardization of observation instrument usage. Standardization of measurements made using these instruments is considered to be achieved when the following three conditions are met: (a) when behaviors observed are classified the same way by different persons using an instrument's categories, (b) when measurements of the behaviors classified are made using the same metric so that scalar identity is achieved across occasions in which the

instrument is used, and (c) when systematic measurement errors do not occur.

The first condition requires that a "standard language" be shared and used by different observers to classify events they observe. The second condition means that when an instrument is used to make measurements of behaviors occurring in different locations, cultures, or at different times, the measurements obtained will represent quantitatively identical scales (Poortinga, 1975). That is, differences in the measurements made will represent actual differences in the extent to which behaviors observed occurred, while equal measurements will indicate equal magnitudes of behaviors observed (within limits imposed by random errors of measurement). Such scalar identity signifies not only that an instrument measures the same attributes in different cultures, but that the same quantitative scale is used in each culture to measure those attributes.[2] The third condition requires that biases will not affect measurements made, such that those measurements systematically differ from the hypothetical "true values" of the behaviors observed (Schumacher, 1981). This means, for example, that observers will not make different measurements due to differing sensitivities to behavioral subtleties in one or more of the cultures studied (Longabaugh, 1980, p. 106; Moore, 1969, p. 255; Schweizer, 1978, pp. 134-135), and that time unit distortions, which are sometimes associated with use of the interval recording method, will not occur (Pfau, 1981).

Past Usage of Category Systems

A number of researchers have used category systems to measure and compare naturally occurring behaviors in different cultures and nations.[3]

Investigations in which these instruments were used to compare behaviors occurring in different countries have included studies of parent and child behaviors in Japan and the U.S.A. (Caudill & Weinstein, 1969; Caudill & Frost, 1974) and in Yugoslavia and the U.S.A. (Lewis & Ban, 1977); differences in infant separation protest in Guatemala and in the U.S.A. (Lester et al., 1974); and child-holding patterns in different societies (Richards & Finger, 1975). A number of researchers have also studied similarities and differences in the classroom behaviors of teachers and students in different countries. Tisher (1970) compared Australian, U.S., and New Zealand teacher behaviors, while other studies compared J.S. teaching with that occurring in Great Britain (Birrell, 1974), in the Bahamas (Ray & Ray, 1976), and in the Kingdom of Nepal (Pfau, 1977). Category systems have also been used to study the social patterns of urban pedestrians in Middle Eastern and Western countries (Berkowitz, 1971), sexual differences in methods of carrying books by students in several Central American and North American countries (Jenni, 1976), and nonverbal behaviors during conversations in Germany, Italy, and the U.S.A. (Shuter, 1977).

Studies conducted within single countries of different cultural and subcultural groups, in which category systems were used, have included comparisons of proxemic behaviors during conversations of Arab and American students (Watson & Graves, 1966), of Anglo-, Black-, and Mexican-Americans observing animals at a U.S. zoo (Baxter, 1970), and during interactions of black, Puerto Rican, and white student dyads on school playgrounds around New York city (Aiello & Jones, 1971). In addition, interactions between members of nine different ethnic groups at the University of Guam have been studied (Brislin, 1971), as have nonverbal behaviors of Protestant Americans of Anglo-Saxon descent and of American Jews (Shuter, 1979), the behaviors of mothers and children from different social classes and cultural groups in Israel (Greenbaum & Landau, 1977) and in the U.S.A. (Tulkin & Cohler, 1973; Tulkin, 1977; and Moss & Jones, 1977), and classroom behaviors in Amish and non-Amish schools in the U.S.A. (Payne, 1970).

Studies in which category systems have been used to study the behaviors of a single cultural group within a single country have been even more numerous.

Variations of category systems have also been used in several significant cross-cultural studies. The most extensive and influential of these was the "Six Cultures Study", in which child-rearing and child behaviors in different cultures were described and compared (Whiting & Whiting, 1975). The approach used included having observers write

extensive running accounts (called protocols) of behaviors which occurred and the contexts of those behaviors, rewriting the acccunts in more and clearer detail as soon as possible, and then later coding the descriptions using a number of behavioral categories. Similar approaches, in which the running accounts have sometimes been called "specimen records", have included studies of behaviors occurring in an American and an English town (Barker & Barker, 1963, 1978; Schoggen, Barker, & Barker, 1963, 1978), and of child behaviors in Japan and in the U.S.A. (Caudill & Schooler, 1973).[4]

Another variation of category systems, the checklist, was used to study and compare science teaching in Britain and in Canada (Hacker, Hawkes, & Heffernan, 1979).

A major accomplishment of these studies has been to demonstrate the range of theoretical and heuristic concerns, and the diversity of cultural and behavioral situations to which the systematic study of naturally occurring behavior is applicable. They have stimulated thinking about what can be done, and provided a basis upon which future work and thinking can build.

For the most part, however, these studies represent only an incomplete beginning to the standardization of behavioral measurements across cultures, nations, and time, for reasons to be now discussed.

Limitations of Past Category System Usage

A researcher studying behaviors occurring in different cultures, or in the same culture at different times, may wish to compare existing data gathered by others who used a particular category system. Or, he or she may conduct or coordinate studies aimed at measuring behaviors in cultures of interest, and then compare the measurements obtained. A major consideration of such researchers should be the degree to which the measurements compared are standardized.

Table 2 indicates some ways of trying to help ensure that standardization of category system and other observation instrument usage is achieved across cultures. These ways range from what this writer and others consider to be a "rigorous approach" (Longabaugh, 1980, pp. 104 & 106; Brislin, 1980, pp. 408-409; Campbell, 1970, pp. 70-71), to much more questionable approaches for measuring and comparing behaviors. As can be seen by looking at Table 2, the most rigorous approach uses observers from each culture studied to help determine if standardized measurements are made in those cultures. Using observers with such diverse backgrounds increases the chance that differences in the way behaviors are classified using an instrument, systematic measurement errors that may occur, and differences in the scalar identity of measurements made in each culture will be detected.

Using the approaches indicated in Table 2 to classify the cross-cultural studies of behavior mentioned before yields the results shown in Table 3. As can be seen, nearly all of

TABLE 2

EXAMPLES OF TECHNIQUES FOR HELPING STANDARDIZE MEASUREMENTS MADE USING CATEGORY SYSTEMS IN DIFFERENT CULTURES

A. A Rigorous Approach[a]

Observers from Culture A and Culture(s) B (C,D,...) observe behaviors of Culture A using a category system and reach high agreement among measurements made. These same observers, without prior discussion, observe behaviors of Culture(s) B (C,D,...) using the category system and again reach high agreement among measurements made. Measurements of behaviors made by these observers using the category system are compared between the cultures.

B. Semi-Rigorous Approaches

Approach B_1: Observers from Culture A observe behaviors of Culture A using a category system and reach high levels of agreement among measurements made. One or more of these observers goes to Culture(s) B (C,D,...) and either makes measurements directly using the category system or trains persons from Culture(s) B (C,D,...) to use the category system in that culture until their measurements agree highly with those made by the Culture A observer(s). Measurements made of Culture A, B (C,D,...) behaviors using the category system are then compared.

Approach B_2: Similar to Approach B_1 except that observers from Culture(s) B (C,D,...) take the initiative in learning to use a category system developed in Culture A. The Culture B (C,D,...) observers reach high levels of agreement with Culture A observers when observing Culture A behaviors. Measurements made by Culture A observers in Culture A are compared with measurements made by Culture B (C,D,...) observers in Culture(s) B (C,D,...).

Approach B_3: One or more observers use a category system at or near a single location in one country to measure the behaviors of two or more cultural groups. Formal checks indicate that high agreement or stability is achieved between measurements of the same events. Observer backgrounds are similar to some but not all of the cultures observed, or represent cultures different from those observed.

TABLE 2 (Continued)

C. Approaches of More Questionable Rigor

Approach C_1: An observer or observers attempt to use a category system the same way as others have previously used it, by reading descriptive materials or by learning to use it from a previous investigator. Although checks of agreement may be made among observers using the instrument in the new study or sub-study, formal checks are not made to determine if use of the instrument in the new study is similar to previous usage in other cultures. However, measurements made in the different cultures and studies are compared.

Approach C_2: Observers reach understandings among themselves about how a category system is to be used. These observers then use the instrument in Cultures A, B, (C,D,...) and compare measurements made in these cultures. Formal checks of agreement are not made, however, to determine if these observers agree highly among themselves when observing and describing the same events.

Approach C_3: An observer uses the same category system to make observations in two or more cultures. Formal checks of agreement and stability of usage are not made, however.

D. Highly Questionable Approach

Measurements are compared of behaviors having the same general label but which were measured by different instruments used by different investigators. No formal checks of agreement are made to determine the equivalence of measurements of the same events resulting from use of the different instruments, although checks of agreement may be made between observers using a particular instrument in any one of the investigations which generated measurements.

---

[a]In this and other approaches described, standardization of usage also requires that systematic measurement errors such as "time unit distortion" be controlled and eliminated.

TABLE 3

CLASSIFICATION OF CROSS-CULTURAL STUDIES OF BEHAVIOR WITH RESPECT TO THE RIGOR OF PROCEDURES USED TO STANDARDIZE MEASUREMENTS COMPARED

A. Rigorous Approach

(no studies were identified which used such an approach)

B. Semi-Rigorous Approach

Approach B_1: Caudill and Weinstein (1969)
Berkowitz (1971)
Caudill and Frost (1974)
Lester et al. (1974)
Ray and Ray (1976)
Pfau (1977)
Shuter (1977)

Approach B_2: Hacker, Hawkes, and Heffernan (1979)

Approach B_3: Watson and Graves (1966)
Baxter (1970)
Payne (1970)
Aiello and Jones (1971)
Brislin (1971)
Tulkin and Cohler (1973)
Greenbaum and Landau (1977)
Moss and Jones (1977)
Tulkin (1977)
Shuter (1979)

TABLE 3 (Continued)

C. Approaches of More Questionable Rigor

Approach C_1: Tisher (1970)

Approach C_2: Barker and Barker (1963, 1978)

Schoggen, Barker, and Barker (1963, 1978)

Caudill and Schooler (1973)[a]

Whiting and Whiting (1975)

Approach C_3: Birrell (1974)

Richards and Finger (1975)[b]

Jenni (1976)

Lewis and Ban (1977)

D. Highly Questionable Approach

Konner (1977, pp. 294-295)[c]

Minge-Klevana (1980)[c]

[a]Agreement checks in this study dealt with only the second step of the specimen record procedure used (i.e., agreement between codings of the same specimen records made by different observers) but did not deal with the first step (i.e., the degree to which specimen records of the same events made by different observers were similar).

[b]Although observer agreement checks were not reported in this study, the behaviors classified were so obvious that some persons may consider this study to represent a semi-rigorous approach.

[c]These authors were aware of inadequacies in the data compared.

the studies reviewed have used "semi-rigorous" or "more questionable" approaches to help standardize the measurements compared.

A major problem of the "semi-rigorous approaches is that observer drift may have occurred when the observation instrument was used to make measurements of different cultural groups (Kazdan, 1977; Longabaugh, 1980, pp. 107-109). That is, those who conducted studies using these approaches assumed that an observer or observers who used an observation instrument in standard ways when observing members of one culture, transferred standardized usage of the instrument to other cultures when measurements were made. This is an untested assumption of these studies -- and can be viewed as a limitation of them and of "semi-rigorous" approaches in general.

The procedures used in the "more questionable" studies, besides not controlling for observer drift in usage, led to comparisons being made of data whose precision, as indicated by tests of observer agreement, is unknown. This means that the scalar identity of measurements made in these studies is open to even more question than those of the "semi-rigorous" approaches -- and this is considered to be a serious limitation of these "more questionable" studies.[5]

However, perhaps an even greater limitation of nearly all of the studies reviewed is that almost none have established a sufficient basis so that future researchers who may wish to gather and compare data with these past studies can ensure that their use of the "same" instrument is indeed the same. That

is, almost none of the investigators who conducted these studies has provided or otherwise retained enough information to permit standardization of observation instrument usage to be achieved between these past studies and future studies which these researchers or others may wish to conduct. This means that, in most cases, the scalar identity of measurements made during these past studies and during future studies cannot be estimated, nor can many systematic measurement errors which may have occurred in these studies be detected. As a result, comparisons of measurements made in the future with those made in most of these past studies will be hazardous.

A Suggested Approach

Techniques for helping overcome the problems of standardizing measurements made using category systems have already been indicated in Table 2. That is, one of the more rigorous approaches described in that table can be used to help ensure that observers are making standardized measurements of different cultures at approximately the same time. However, the procedures indicated do not help to ensure that observers in the future will use an observation instrument as it was used in the past. This is so because even expert observers may modify their use of an instrument over time. They may also die. Observer drift in usage over time which may result needs to be controlled if standardized measurements are to be made by observers at different times, either in the same or in different studies.[6]

A way of overcoming this time-related problem as well as the cross-cultural problems discussed before is the following:

1. Preserve Samples of Behavior Observed

Instrument developers and users would, according to this proposed approach, preserve samples of behaviors measured on movie film, audio-video tape recordings, audio recordings, or photographs -- the exact media used being dependent upon the types of behavior measured.[7] Relevant contextual information which is not evident from the preserved behavioral recordings should also be described in sufficient detail so that future observers will have enough information to accurately code the behaviors preserved when they use the instrument.

2. Make Standard Codings of These Preserved Behaviors

The preserved samples of behavior should then be coded by an "expert" observer or by typical observers who participated in the study whose instrument usage is being preserved for future reference. The codings made will constitute a set of preestablished standards against which future measurements can be compared.

3. Prepare Instrument Descriptions

Sufficient information about the instrument used, including other sets of preserved behavior samples and associated "standard codings", should be prepared so that future users can train themselves and others to use the instrument the same way it was used in the past ("Where do . . . ," 1965; Thiagarajan, 1973). These materials should be made available for future use by others.[8]

4. Future Users Test Their Usage

The materials prepared would then be used to train

new observers to make standardized measurements using the instrument and to help detect and correct observer drift from standardized usage which may occur in a study. For example, after training, observers would code previously unseen samples of the preserved behavioral records and their measurements would be compared with the preestablished standard measurements of those records. If high agreement is reached, this will indicate that the new observers are making measurements in a standardized manner. After high agreement is reached, additional checks using the preserved materials can be made from time to time to help detect and correct observer drift from standardized usage which may occur (see Roebuck, Aspy, Sadler, & Willson, 1974, to see how this has been done in the past).

5. Determine Standardization Across Cultures or Time

The techniques outlined in Table 2 could then be used to help standardize measurements made across cultures and longer periods of time.

For example, the procedures described above could be used with the Rigorous Approach described in Table 2 by having observers from Cultures A and B (C, D, . . .) receive training until their measurements of the preserved behavioral recordings agree highly with the preestablished standard codings originally prepared. These observers would then jointly observe behaviors in Cultures A, B (C, D, . . .) and determine if measurements of the same behavioral events they make in those cultures also agree highly. If

agreement is reached, then standardization of measurements made in those cultures is indicated, and comparisons of those measurements can be made.

Similarly, measurements made at one time (t_1) can be compared with measurements made using the same observation instrument in another culture or in the "same" culture at a future time (t_2). This could be done by having observers receive training using preserved behavioral records and training materials prepared when the instrument was used at the earlier time (t_1). After observers reach high agreement with the preestablished standards associated with the training materials, measurements could be made at the future time (t_2) and compared with those made at the earlier time (t_1). This technique, if followed, represents Approach B_2 of Table 2, where Culture A is the behavioral situation measured at time t_1 and Culture B is either a quite different culture at time t_2 or a variation of Culture A which has evolved over time.

Some Needed Research and Thinking

Although the suggestions made in the previous sections provide a framework for discussion and action, additional information and thinking are needed if standardization is to be achieved with confidence. For example, information helpful to persons wishing to standardize behavioral measurements across cultures and time would be provided by research which answered the following questions.

1. Can an observer who uses an observation instrument

in a standard way in one culture, transfer standardized usage of the instrument to another culture (or language) when he or she observes behaviors in the second culture or trains others to do so? An assumption of the widely used "Semi-Rigorous Approaches" described in Table 2 is that, yes, such transferability of standardized usage can and does occur. As indicated before, this assumption is yet to be tested.

2. What techniques should investigators use to achieve standardization of observation instrument usage across cultures? The "Rigorous Approach" is one possible procedure. Is this approach sufficient? Is it too rigorous? Are other approaches more practical and satisfactory?

3. What should be done if observers differ in the measurements they make in different cultures? Caudill, for instance, found that measurements made by observers in Japan and in the U.S.A. differed somewhat from his own (his being the standard against which theirs were judged). In order to make the measurements of these observers more equivalent, Caudill used a "weighing" procedure to adjust their scores. Are such weighing procedures a promising approach to use when differences in instrument usage are found to occur across cultures? (See Caudill and Weinstein, 1969, pp. 24-25, and Caudill and Frost, 1973, p. 7, for details).

4. Does the two-step procedure involved in measuring behavior by first writing specimen records or protocols and

then coding these written descriptions, result in measurements which are accurate enough to compare across cultures? There are some indications that such may not be the case (Spain and Hollenbeck, 1975; Levine, 1977). The extent to which this procedure can be standardized across cultures needs to be studied more, given the fairly widespread use of such descriptions for comparative purposes.

In addition, several other areas related to the use of category systems in comparative studies are in need of thought and investigation. These include the questions of (a) how category systems should be developed or modified to best ensure that they are suitable for the comparative purposes for which they are to be used, and (b) what additional techniques should be used and what additional information should be gathered so that the measurements provided by category system usage can be validly interpreted beyond the specific events quantified (since category systems, by themselves, do not provide much of a basis for understanding and explaining the events measured). Although some information exists concerning these areas (ex., Goodenough, 1970, chap. 4; Pfau, 1981, pp. 31-34), more is needed.

It is hoped that efforts will soon be made to answer these questions and others which this monograph will surely raise. It is also hoped that the suggestions made in this essay will provide a useful guide for action until standards for measuring behavior using direct observation techniques are more formally established by a consensus of concerned scholars (Standardization Basics, 1977a, 1977b).

Footnotes

I would like to thank Richard W. Brislin and Carmi Schooler for commenting upon an earlier draft of this monograph which is based upon a paper presented at the annual meeting of the Society for Cross-Cultural Research, Syracuse, New York, February 1981.

[1]A visit during July 1981 to the National Bureau of Standards reference collections at Gathersburg, Maryland, inquiries to the American National Standards Institute (ANSI), to the American Society for Testing and Materials (ASTM), and to the International Organization for Standardization (ISO), and a review of related literature yielded no standards for measuring the occurrence of human or animal behavior except for some concerning psychological testing.

[2]Scalar identity is considered necessary if the scores of culturally different groups are to be compared, according to Davidson (1977, p. 50).

[3]The author welcomes information from readers about other cross-cultural studies which used category systems to measure naturally occurring behaviors in non-experimental settings.

[4]Although the two-step (i.e., making a written description which is then coded) specimen record and protocol approaches provide a great deal of rich contextual information about behaviors observed, they seem to result in measurements which are not as accurate as those made by more

directly coding behavior observed (Spain and Hollenbeck, 1975; Levine, 1977). Checklists also seem to yield measurements which are less precise than those provided by other kinds of category systems, are more prone to distorting estimates of the extent to which behaviors occur, and do not lend themselves as well to the study of sequences of events (Dunkin and Biddle, 1974, p. 71). As a result, these variations do not seem to lend themselves as well to standardizing measurements across cultures, and are differentiated from other kinds of category systems in this article for that reason.

[5]This does not mean that scalar identity may not have been approximated in some cases, nor that a great deal of thought provoking and useful data was not gathered by many of these studies. However, the degree to which measurements made in these studies were standardized is open to question.

[6]Such observer drift is an example of what Campbell and Stanley (1966) call "instrument decay".

[7]For example, if the locations of persons in a room are being studied, photographs may be sufficient. If an analysis of verbal behavior is being conducted, audio tape recordings may suffice. It should be noted though that specimen records, transcripts, or other kinds of narrative descriptions are not considered to be suitable for the preservation of realia as required by this step.

[8]See Herbert and Attridge, 1975, for guidelines about what to include in such training materials.

References

Aiello, J. R., & Jones, S. E. Field study of the proxemic behavior of young school children in three subcultural groups. Journal of Personality and Social Psychology, 1971, 19, 351-356.

Altmann, J. Observational study of behavior: Sampling methods. Behaviour, 1974, 49, 227-267.

Barker, R. G., & Barker, L. S. Social actions in the behavior streams of American and English children. In R. G. Barker (Ed.), The stream of behavior: Explorations of its structure & content. New York: Appleton-Century-Crofts, 1963.

Barker, R. G., & Barker, L. Social actions of American and English children and adults. In R. G. Barker & Associates, Habitats, environments, and human behavior. San Francisco: Jossey-Bass, 1978.

Baxter, J. C. Interpersonal spacing in natural settings. Sociometry, 1970, 33, 444-456.

Berkowitz, W. R. A cross-national comparison of some social patterns of urban pedestrians. Journal of Cross-Cultural Psychology, 1971, 2, 129-144.

Birrell, H. V. A comparative study of the verbal behavior of teachers in open education classrooms in England and the United States (Doctoral dissertation, Syracuse

University, 1974). Dissertation Abstracts International, 1976, 36, 6609A-6610A. (University Microfilms No. 76-7883)

Brislin, R. W. Interaction among members of nine ethnic groups and belief-similarity hypothesis. Journal of Social Psychology, 1971, 85, 171-179.

Brislin, R. W. Translation and content analysis of oral and written materials. In H. C. Triandis & J. W. Berry (Eds.), Handbook of cross-cultural psychology (Vol. 2). Boston: Allyn & Bacon, 1980.

Campbell, D. T. Natural selection as an epistemological model. In R. Naroll & R. Cohen (Eds.), A handbook of method in cultural anthropology. New York: Natural History Press, 1970.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Caudill, W., & Frost, L. A comparison of maternal care and infant behavior in Japanese-American, American, and Japanese families. In W. P. Lebra (Ed.), Youth, socialization, and mental health. Honolulu: University Press of Hawaii, 1974.

Caudill, W. A., & Schooler, C. Child behavior and child rearing in Japan and the United States: An interim report. Journal of Nervous and Mental Disease, 1973, 157, 323-338.

Caudill, W., & Weinstein, H. Maternal care and infant behavior in Japan and America. Psychiatry, 1969, 32 12-43.

Davidson, A. R. The etic-emic dilemma: Can methodology provide a solution in the absence of theory? In Y. P. Poortinga (Ed.), Basic problems in cross-cultural psychology. Amsterdam: Swets & Zeitlinger, 1977.

Dunkin, M. J., & Biddle, B. J. The study of teaching. New York: Holt, Rinehart, & Winston, 1974.

Goodenough, W. H. Description and comparison in cultural anthropology. Chicago: Aldine, 1970.

Greenbaum, C. W., & Landau, R. Mothers' speech and the early development of vocal behavior: Findings from a cross-cultural observation study in Israel. In P. H. Leiderman, S. R. Tulkin, & A. Rosenfeld (Eds.), Culture and infancy: Variations in the human experience. New York: Academic Press, 1977.

Hacker, R. G., Hawkes, R. L., & Heffernan, M. K. A cross-cultural study of science classroom interactions. British Journal of Educational Psychology, 1979, 49, 51-59.

Herbert, J., & Attridge, C. A guide for developers and users of observation systems and manuals. American Educational Research Journal, 1975, 12, 1-20.

Jackson, D. A., Della-Piana, G. M., & Sloane, H. N. How to establish a behavior observation system. Englewood Cliffs: Educational Technology Publications, 1975.

Jenni, M. A. Sex differences in carrying behavior. Perceptual and Motor Skills, 1976, 43, 323-330.

Johnson, A. W. Quantification in cultural anthropology. Stanford: Stanford University Press, 1978.

Kazdin, A. E. Artifact, bias, and complexity of assessment: The ABCs of reliability. Journal of Applied Behavior Analysis, 1977, 10, 141-150.

Konner, M. Infancy among the Kalahari Desert San. In P. H. Leiderman, S. R. Tulkin, & A. Rosenfeld (Eds.), Culture and infancy: Variations in the human experience. New York: Academic Press, 1977.

Lester, B. M., Kotelchuck, M., Spelke, E., Sellers, M. J., & Klein, R. E. Separation protest in Guatemalan infants: Cross-cultural and cognitive findings. Developmental Psychology, 1974, 10, 79-85.

Levine, A. J. Naturalistic observation: Validity of frequency data. Psychological Reports, 1977, 40, 1311-1338.

Lewis, M., & Ban P. Variance and invariance in the mother-infant interaction: A cross-cultural study. In P. H. Leiderman, S. R. Tulkin, & A. Rosenfeld (Eds.), Culture and infancy: Variations in the human experience. New York: Academic Press, 1977.

Longabaugh, R. The systematic observation of behavior in naturalistic settings. In H. C. Triandis & J. W. Berry (Eds.), Handbook of cross-cultural psychology (Vol. 2). Boston: Allyn & Bacon, 1980.

Minge-Klevana, W. Does labor time decrease with industrialization? A survey of time-allocation studies. Current Anthropology, 1980, 21, 279-298.

Moles, J. A. Standardization and measurement in cultural anthropology: A neglected area. Current Anthropology, 1977, 18, 235-258.

Moore, F. W. Codes and coding. Behavior Science Notes. 1969, 4, 247-263.

Moss, H. A., & Jones, S. J. Relations between maternal attitudes and maternal behavior as a function of social class. In P. H. Leiderman, S. R. Tulkin, & A. Rosenfeld (Eds.), Culture and infancy: Variations in the human experience. New York: Academic Press, 1977.

Nunnally, J. C. Psychometric Theory (2nd ed.). New York: McGraw-Hill, 1978.

Payne, J. I. Analysis of teacher-student classroom interaction in Amish and non-Amish schools using the Flanders Interaction Analysis technique (Doctoral dissertation, Pennsylvania State University, 1970). Dissertation Abstracts International, 1971, 32, 251A. (University Microfilms No. 71-16,650)

Pfau, R. H. The cross-national comparison of classroom behaviors. Paper presented at the annual meeting of the Comparative and International Education Society, Toronto, February 1976. (ERIC Document Reproduction Service No. ED 120 260)

Pfau, R. H. A cross-national comparison of classroom behaviors: Based upon a survey conducted within Nepal using the Flanders Interaction Analysis Categories and Caldwell's Activity Categories Instrument (Doctoral

dissertation, University of Pittsburgh, 1977). Dissertation Abstracts International, 1978, 38, 7034A. (University Microfilms No. 78-09,602)

Pfau, R. H. The comparative study of classroom behaviors. Comparative Education Review, 1980, 24, 400-414.

Pfau, R. H. Sources of invalidity when comparing classroom behaviors across cultures and nations. Paper presented at the annual meeting of the Comparative and International Education Society, Tallahassee, March 1981. (ERIC Document Reproduction Service No. ED 202 134)

Poortinga, Y. H. Some implications of three different approaches to intercultural comparison. In J. W. Berry & W. J. Lonner (Eds.), Applied cross-cultural psychology. Amsterdam: Swets & Zeitlinger, 1975.

Ray, R. D., & Ray, M. R. A systems approach to behavior II: The ecological description and analysis of human behavior dynamics. Psychological Record, 1976, 26, 147-180.

Richards, J. L., & Finger, S. Mother-child holding patterns: A cross-cultural photographic survey. Child Development, 1975, 46, 1001-1004.

Roebuck, F. N., Aspy, D. N., Sadler, L. L., & Willson, M. A. Maintaining reliability in a longitudinal study. Monroe: National Consortium for Humanizing Education, Northeast Louisana University, 1974. (ERIC Document Reproduction Service No. ED 106 730)

Schoggen, M., Barker, L. S., & Barker, R. G. Structure of the behavior of American and English children. In R. G.

Barker (Ed.), The stream of behavior: Explorations of its structure & content. New York: Appleton-Century-Crofts, 1963.

Schoggen, M. F., Barker, L. S., & Barker, R. G. Behavior episodes of American and English children. In R. G. Barker and Associates, Habitats, environments, and human behavior. San Francisco: Jossey-Bass, 1978.

Schumacher, R. B. F. Systematic measurement errors. Journal of Quality Technology, 1981, 13, 10-24.

Schweizer, T. Data quality and data quality control in cross-cultural studies. Behavior Science Research, 1978, 13, 125-150.

Shuter, R. A. A field study of nonverbal communication in Germany, Italy, and the United States. Communication Monographs, 1977, 44, 298-305.

Shuter, R. A study of nonverbal communication among Jews and Protestants. Journal of Social Psychology, 1979, 109, 31-41.

Spain, D. H., & Hollenbeck, A. R. Coding non-Western behavior stream data: A problem of reliability. Journal of Social Psychology, 1975, 97, 137-138.

Standardization basics: Part I, the what and why of standards. ASTM Standardization News, 1977a, 5(2), 24-27.

Standardization basics: Part II, the people, process, and principles of standards development. ASTM Standardization News, 1977b, 5(3), pp. 21-25; 60.

Thiagarajan, S. Instructional systems for interactional

systems. Classroom Interaction Newsletter, 1973, 9(1), 13-22.

Tisher, R. P. A study of verbal interaction in science classes and its association with pupils' understanding in science. St. Lucia: University of Queensland Press, 1970.

Triandis, H. C. Presidential address: Impediments to the progress of cross-cultural psychology. In Y. H. Poortinga (Ed.), Basic problems in cross-cultural psychology. Amsterdam: Swets & Zeitlinger, 1977.

Tulkin, S. R. Social class differences in maternal and infant behavior. In P. H. Leiderman, S. R. Tulkin, & A. Rosenfeld (Eds.), Culture and infancy: Variations in the human experience. New York: Academic Press, 1977.

Tulkin, S. R., & Cohler, B. J. Childrearing attitudes and mother-child interaction in the first year of life. Merrill-Palmer Quarterly, 1973, 19, 95-106.

Watson, O. M., & Graves, T. D. Quantitative research in proxemic behavior. American Anthropologist, 1966, 68, 971-985.

Where do we go from here? Classroom Interaction Newsletter, 1965, 1(1), 5.

Whiting, B. B., & Whiting, J. W. M. Children of six cultures: A psycho-cultural analysis. Cambridge, Mass.: Harvard University Press, 1975.